



Article Development of Monthly Reference Evapotranspiration Machine Learning Models and Mapping of Pakistan—A Comparative Study

Jizhang Wang ¹, Ali Raza ¹, Yongguang Hu ^{1,*}, Noman Ali Buttar ^{1,2}, Muhammad Shoaib ³, Kouadri Saber ⁴, Pingping Li ⁵, Ahmed Elbeltagi ⁶ and Ram L. Ray ⁷

- ¹ School of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China; whxh@ujs.edu.cn (J.W.); alir3201@gmail.com (A.R.); noman_buttar@yahoo.com (N.A.B.)
- ² Department of Agricultural Engineering, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan
- ³ Department of Agricultural Engineering, Bahauddin Zakariya University, Multan 60000, Pakistan; muhammadshoaib@bzu.edu.pk
- ⁴ Laboratory of Water and Environment Engineering in Saharan Environment, Department of Civil and Hydraulic Engineering, Faculty of Applied Sciences, University of Kasdi Merbah-Ouargla, PB 147 RP, Ouargla 30000, Algeria; kouadri.saber@univ-ouargla.dz
- ⁵ School of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China; lipingping@ujs.edu.cn
- ⁶ Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt; ahmedelbeltagy81@mans.edu.eg
- ⁷ Department of Agriculture, Nutrition and Human Ecology, College of Agriculture and Human Sciences, Prairie View A&M University, Prairie View, TX 77446, USA; raray@pvamu.edu
- Correspondence: deerhu@ujs.edu.cn; Tel.: +86-138-1515-1176

Abstract: Accurate estimation of reference evapotranspiration (ET_o) plays a vital role in irrigation and water resource planning. The Penman-Monteith method recommended by the Food and Agriculture Organization (FAO PM56) is widely used and considered a standard to calculate ET₀. However, FAO PM56 cannot be used with limited meteorological variables, so it is compulsory to choose an alternative model for ET_o estimation, which requires fewer variables. This study built ten machine learning (ML) models based on multi-function, neural network, and tree-based structure against the FAO PM56 method. For this purpose, time series temperature data on a monthly scale are only used to train ML models. The developed ML models were applied to estimate ET₀ at different test stations and the obtained results were compared with the FAO PM56 method to verify and validate their performance in ET_o estimation for the selected stations. In addition, multiple statistical indicators, including root-mean-square error (RMSE), coefficient of determination (\mathbb{R}^2), mean absolute error (MAE), Nash–Sutcliffe efficiency (NSE), and correlation coefficient (r) were calculated to compare the performance of each ML model on ET_o estimation. Among the applied ML models, the ET_o tree boost (TB) ML model outperformed the other ML models in estimating ET₀ in diverse climatic conditions based on statistical indicators (R², NSE, r, RMSE, and MAE). Moreover, the observed R², NSE, and r were the highest for the TB ML model, while RMSE and MAE were found to be the lowest at the study sites compared to other applied ML models. Lastly, ET_o point data yielded from the TB ML model was used in an interpolation process to create monthly and annual ETo maps. Based on the ET_o maps, this study suggests mainly a focus on areas with high ET_o values and proper irrigation scheduling of crops to ensure water sustainability.

Keywords: reference evapotranspiration; machine learning models; neural networks-based models; tree-based models; multifunction-based models; climatic regions



Citation: Wang, J.; Raza, A.; Hu, Y.; Buttar, N.A.; Shoaib, M.; Saber, K.; Li, P.; Elbeltagi, A.; Ray, R.L. Development of Monthly Reference Evapotranspiration Machine Learning Models and Mapping of Pakistan—A Comparative Study. *Water* 2022, *14*, 1666. https://doi.org/10.3390/w14101666

Academic Editors: Hakan Başağaoğlu, Debaditya Chakraborty and Marcio Giacomoni

Received: 31 March 2022 Accepted: 19 May 2022 Published: 23 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

Water scarcity has become a global issue and a challenging task for arid and semiarid areas. Because of the rapid growth of the world's population, each country attempts to utilize water with better management by adopting innovative ways. [1]. Upgrading water management systems and quantitative irrigation scheduling are essential to solving water shortages. The proper amount of water given at the right time to crops is called irrigation scheduling which may be computed in terms of evapotranspiration (ET). It is a combination of evaporation and transpiration processes [2]. The determination of ET is directly quantified by using the lysimeter, eddy covariance, and Bowen ratio energy balance methods. ET estimation with these methods involves high capital, operation, and maintenance costs, which may limit their use in practice [3–5]. As a result, relying on the indirect technique is the best choice for quantification. One of these indirect techniques is using empirical models that can be classified into multiple classes, including temperaturebased, radiation-based, mass transfer-based models, etc. The reference evapotranspiration (ET_{o}) is estimated in almost all empirical models. The reason for this is that calculating ET for each crop is challenging. As a result, ET_o is first estimated using indirect techniques, and crop coefficients are then utilized to predict ET of each desired crop. Several attempts have been made to estimate ET_o correctly. However, the Food and Agriculture Organization (FAO) of the United Nations (UN) and the American Society of Civil Engineers (ASCE) ET committee recommend the Penman-Monteith (FAO PM56) method, which Allen introduced in 1998 and validated in various climatic conditions [4,6].

Many sites across the globe do not have a full set of meteorological data for FAO PM56 ET_o calculations. However, recent studies reported ML-based predictions of the FAO PM56 ET_0 using a full set of meteorological variables [7,8], which also unveiled the relative importance and interdependencies and interrelations among the variables. In a different study, ET_o was computed using FAO PM56 with a full set of meteorological variables, and the results were compared against predictions by a deep neural network (NN) model that uses solar radiation (R_n) as a sole meteorological predictor [9]. R_n was found to be the most critical meteorological variable in calculating ET_0 in a semi-arid region by Basagaoglu et al. [7]. However, sunshine hours (N) can be used as a surrogate variable for R_s when the latter is not available or precarious as in the current study. Zhou et al. [10] also endorsed this statement as they explored the potential of deep factorization machine (DeepFM), three-gradient boosting (gradient boosting with categorical features support (CatBoost), light gradient boosting (LightGBM), extreme gradient boosting (XGBoost)), three tree-based (gradient boosting decision tree (GBDT), random forest (RF), decision tree (DT)), and one support vector machine (SVM-RBF) machine learning (ML) models for modeling daily ET_0 in China by considering climatic variables (N, temperature (T), relative humidity (RH), and wind speed (U)) of 12 stations located in humid and arid climatic conditions. Previous studies have indicated that N was more closely related to R_s than other meteorological variables [11–13]. Therefore, this study selected N as a substitute for Rs.

Limited access to all parametric data and discovered ambiguity in the dependability of meteorological parameters or unavailability of climatic variables for numerous sites becomes a significant impediment to ET_o estimation using the FAO PM56 method [14,15]. Although essential climatic variables are now accessible for climatic stations, still, climatic data lack at many locations due to a limited number of automated weather stations deployed at specific sites, and data quality is a concern due to installing obsolete weather stations. The sensor's operation is unreliable and reveals numerous variations that need data calibration to be usable. In this case, ET_o estimates derived using the FAO PM56 method are inconsistent and create a space to develop new methods to reliably estimate ET_o based on fewer inputs and an approach for the use of at least the FAO PM56 method.

Since a full set of meteorological variables is not available at many sites across the globe, numerous studies have been reported that correlate their proposed equations using limited meteorological data with the FAO PM56 method to develop alternative methods for

 ET_o estimation. However, the results of these methods are often found inconsistent [16–19]. However, over the past two decades, ML synchronized with several algorithms were applied for ET_o modeling using limited climatic data [20–28]. The resulting ML models indicate the best ET_o modeling choice leading to the FAO PM56 method. The adaptation in the advantage of ML models is mainly the selection of input variables versus output variables. In addition, adding or removing an input variable using the ML technique to create an output model compared to conventional methods entices researchers in additional hydrological studies to develop the best input–output relationship [29–35].

The recent studies on ET_o modeling highlighted this trending topic and were overwhelming among hydrologists and meteorologists. Yin et al. [36] examined the performance of the artificial neural network (ANN), SVM, and three empirical models for estimating the daily ET_o in a mountainous inland watershed in northwest China using historical weather daily data. They reported SVM performed best in the study region. Wen et al. [37] estimated ET_o of an arid region in China using daily climatic variables (T_{max} , T_{min} , U, R_s) by applying two ML models of ANNs and SVM against three empirical (Hargreaves (Ha), Ritchie (Ri) Priestley, and Taylor (PT)) models. The authors used maximum and minimum temperature as the model input. It was observed that the estimated daily ET_o was reasonable using limited meteorological variables. Wang et al. [38] investigated the performance of two ML models, i.e., gene expression programming (GEP) and ANNs, for estimating daily ET_o using climatic data (T_{max} , T_{min} , U, RH, and N) of four meteorological stations located in the karst region of Guangxi Province in China. The study revealed that GEP based on fewer climatic inputs could produce simple explicit mathematical formulas that are easier to use than the applied ANN models.

Sanikhani et al. [39] used four ML models (multilayer perceptron (MLP), generalized regression neural network (GRNN), radial basis function neural network (RBFNN), and adoptive neuro fuzzy interference system (ANFIS)) to predict the monthly ET_o using three climatic variables (T_{max} , T_{min} and extraterrestrial radiation (R_a)) from two stations in Turkey. They developed the ML models based on temperature data, and the accuracy was examined using the Hargreaves–Samani (HS) empirical equation. The GEP and GRNN in Antalya outperformed the RBFNN and ANFIS in Isparta. Mohammadrezapour et al. [40] applied three ML models (SVM, GEP and ANFIS) using different input climatic combinations for ET_o modeling. Their results showed that SVM exhibited the best performance among the other ML models. Wu et al. [41] investigated the performance of eight ML models (ANN, RF, GBDT, XGBoost, multivariate adaptive regression spline (MARS), SVM, extreme learning machine (ELM), and kernel-based nonlinear extension of arps decline models (KNEA)) to estimate ET_o using temperature data alone and concluded that tree base ML models of RF, XGBoost, GBDT were outperformed.

Saggi and Jain [42] examined four ML models (deep learning (DL), generalized linear model (GLM), GBDT, and RF) to predict daily ET_0 using climatic variables (T_{max} , T_{min} , U, RH, and N) at Punjab, India. They concluded that the DL model had a better performance than other models. Likewise, Shiri et al. [43] compared a multifunction ML model, i.e., a GEP, with a locally and externally calibrated PT model for ET₀ modeling. They found that the multifunction ML model outperformed the calibrated empirical PT. Tikhamarine et al. [44] calculated (or estimated) monthly ET_o by applying five variants of neural network (NN) models (i.e., embedded grey wolf optimizer (NNGWO), multiverse optimizer (NNMVO), particle swarm optimizer (NNPSO), whale optimization algorithm (NNWOA), and ant lion optimizer (NNALO)) at two stations (Ranichauri, Dar El Beida) in India and Algeria. The ET_o result of the NN models was compared with the Valiantzas (VA) empirical model, and the NNGWO-generated satisfactory output was compared to other NN models. Moreover, Ferreira et al. [45] compared the performance of two ML models (NN, GEP). They indicated that ML models were better than the FAO PM56 method with limited climatic data. In addition, Granata [46] developed four ET_o ML models (regression model tree (M5P), GBDT, RF, and SVM) by using climatic variables obtained from eddy covariance (EC) flux tower stations installed at a site in central Florida characterized by a humid subtropical climate. Based on the number of input variables (R_n , sensible heat flux (H), soil moisture content (W), U, average relative humidity (RH_{avg}) and T_{mean}), three different input variants were created. The results showed that the selected ML models using all three input variants have good capability for ET_o prediction. In addition, M5P proved to be the most accurate, while RF was the least accurate. The M5P ML model based on temperature data performed well and predicted reliable ET_o values. Keshtegar et al. [47] compared the performance of the polynomial chaos expansion and response surface method against the tree and neural network-based ML models for ET_o modeling. Their results showed polynomial chaos expansion and ML models performed better than the empirical models.

Nourani et al. [48] compared ML (MLP, ANFIS, SVR, MLR) and empirical (Hargreaves and Samani (HS), modified Hargreaves and Samani (MHS), Makkink (Mk), and Ri) models for ET_o modeling using monthly meteorological variables (RH, R_s, R_n, R_a, T_{max}, T_{min}, T_{men}, U, and P) from fourteen stations located in different countries (e.g., Turkey, Cyprus, Iraq, Iran, and Libya). The results indicated that ML models outperform traditional empirical models based on different performing indices. Shiri [49] compared a multifunction ML model to six empirical equations (HS, PT, Mk and Turc (Tu), Dalton (Dn) and Trabert (Tr)) for daily ET_0 modeling in a deserted area of Iran using five climatic variables (T_{max} , T_{min} , R_s, U, and RH). The results demonstrated that the calibrated mass transfer-based equations (TU, Dn, Tr) and their corresponding GEP outperformed the other models used in the analyzed sites. This criticized the widely accepted consequence of using temperature-/radiation-based models without data for various sites. Raza et al. [50] reviewed research articles that applied various ML models for ETo modeling using limited climatic data against the FAO PM56 method. The study reviewed recent articles published within the last eight years (2012–2020). This study suggests using ML models with temporal climatic data and creating ET_o maps based on an ML model to present a realistic scenario that is beneficial for crop water distribution.

Although weather records have been readily available in recent years, meteorological data scarcity exists in many sites. In the case of Pakistan (our study area), few weather stations were installed, and climatic data for several locations were found insufficient to calculate ET_0 -based crop water requirements. Thus, the improvement of methods relying on fewer climatic inputs (e.g., temperature data) and the development of ML models for ET_o estimation with minimal climatic records turn into a task of great relevance when standard methods (e.g., FAO PM56) cannot be applied due to excessive input demands or unavailability of climatic parameters, e.g., R_s. For this purpose, ML can be considered one of the best options for developing an ET_0 model. There is no need to calibrate ML against on-site measurements for local calibration. However, the development of an ML model with a known set of input variables against the target variable is a crucial and vital issue that has clearly been addressed in this study. The developed ML models were evaluated at different test stations using only temperature (minimum and maximum) data to validate their performance estimating ET_0 . Finally, ET_0 point data resulting from the best ML model were used in an interpolation process to develop monthly and annual ETo maps, showing the ET_{o} variation over the study region. The ET_{o} maps help agronomists, hydrologists, and agricultural engineers to calculate crop water requirements precisely (e.g., drip and sprinkler irrigation methods) for the cropland to avoid under- and overirrigation problems.

2. Data Collection and Country Profile

The FAO has developed a climatic database called CLIMWAT [51]. It contains geographical parameters (latitude, longitude, altitude) and climatic variables, which include the minimum and maximum temperatures (T_{max} , T_{min}), the average relative humidity (RH_{avg}), wind speed (U), and sunshine hours (N). In this study, the climatic data of 22 stations in Pakistan, were gathered from the CLIMWAT database (Figure 1). Sarfaraz et al. [52] defined Pakistan is located in a temperate zone. The climate is mainly desert, with scorching summers and chilly or frigid winters, with considerable temperature changes at specific sites. Table 1 contains statistical data on accessible climate factors. In addition, as shown in



Figure 2, a boxplot chart was utilized to depict monthly variations of the chosen climatic variables.

Figure 1. Geographical location of the selected climatic stations in Pakistan.

Table 1. Statistical information on the available climatic variables.

Variables	Observations	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Min Temp	264.00	-8.70	30.30	13.78	9.69	-0.18	-0.95
Max Temp	264.00	0.80	44.20	27.55	9.73	-0.57	-0.38
Humidity	264.00	19.00	83.00	49.50	14.18	0.23	-0.55
Wind speed	264.00	17.00	752.00	197.72	145.95	1.33	1.82
Sunshine	264.00	0.80	13.90	7.44	1.87	-0.50	1.72
ETo	264.00	0.63	11.19	4.58	2.39	0.48	-0.31



Figure 2. Boxplot shows monthly variations of the selected climatic variables.

It can also be perceived from Figure 2 that T_{min} increased from January to July and then continuously decreased until December. A similar trend was observed for T_{max} (Figure 2b). This is because of the solar effect from winter to summer seasons (Figure 2e). N observed the maximum during the summer and the minimum when winter started. In addition, RHavg was interconnected with T_{max} and T_{min} because the maximum humidity was observed when the air temperature was high and vice versa. Moreover, RH_{avg} was the highest in the summer season (Jul to Sep) because of the wind effect (Figure 2c). When the wind blows higher in the summer, as observed in Figure 2d, it increases humidity in the air. Overall, the maximum increase in average humidity and the highest humidity value were observed in the summer season.

3. FAO PM56 Method and Development of ML Models

To determine ET_o values, the FAO PM56 standardized method developed by Allen et al. [4] in 1998 is employed. This approach is based on the following equation, which is a mixture of meteorological and aerodynamic parameters:

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma \times \frac{900}{T_{mean} + 273} \times U_2 \times (e_s - e_a)}{\Delta + G(1 + 0.34U_2)}$$
(1)

where:

 ET_o is measured in mm/day;

 R_n = net radiation at the surface (MJ/m²/day);

G = soil heat flux density $(MJ/m^2/day)$;

- T_{mean} = air temperature at 2 m height (°C);
- E_s = saturation vapor pressure (kPa);
- $e_a = actual vapor pressure (kPa);$
- Δ = slope of vapor pressure curve (kPa/°C);
- γ = psychrometric constant (kPa/°C);
- $\gamma = 0.000665 \times p$ (kpa);

 e_s is saturation vapor pressure (kPa), calculated using Equation (2):

$$\mathbf{e}_{\rm s} = \frac{\left(\mathbf{e}_{\rm min} \times \left(\frac{(\rm RH_{\rm max})}{100}\right) + \mathbf{e}_{\rm max} \times \left(\frac{(\rm RH_{\rm min})}{100}\right)\right)}{2} \tag{2}$$

where:

e_{min}: minimum vapor pressure;

e_{max}: maximum vapor pressure;

RH_{min}: minimum relative humidity;

RH_{max}: maximum relative humidity.

 U_2 is wind speed at 2 m height (m/s) calculated using Equation (3):

$$U_2 = \frac{ws \times 4.87 \times 1000}{3600 \times e_{min}(67.8 \times 3 - 5.42)} \tag{3}$$

where:

W_s: wind speed at a specific height above the ground surface;

e_{min}: minimum vapor pressure;

J = 1, 2, 3....n.

CropWat 8.0 given on the official website of the FAO (https://www.fao.org/landwater/databases-and-software/cropwat/en/ (accessed on 10 March 2022)) was used to calculate ET_o values as in the study of Tikhamarine et al. [33]. Moreover, CropWat was developed using the FAO PM56 method, which requires at least five climatic parameters (T_{max} , T_{min} , RH_{avg} , U, N) as well as geographic parameters (longitude, latitude, altitude, periodicity component (*n*)) for ET_o estimation. The statistical values of all the five climatic parameters used to calculate ET_o are included in Table 1. However, the calculated ET_o using the FAO PM56 method (PM ET_o) and CropWat 8.0 are presented in Figure 2f. On the other hand, the average monthly climatic variables and ET_o values labeled as PM ET_o are included in Appendix A.1.

In this study, ten ML models were chosen to estimate ET_{o} . The ML models can be divided into three categories: (i) three tree-based models; (ii) four neural networkbased models; (iii) three multifunction-based models. Training data of only two climatic parameters (T_{max} , T_{min}) were primarily used to train ML models. Afterwards, trained ML models were applied in different climatic stations to calculate monthly ET_{o} using only temperature variables (i.e., T_{max} and T_{min}). The climatic data from 15 stations (15 × 12 = 180 records) were used to train the selected ML models, while climatic data from seven stations (7 \times 12 = 84 records) were used to investigate each ML model's performance. The theoretical details and practical applications of ML models to estimate ET_o are elaborated below.

3.1. Tree-Based ML Models

3.1.1. Single Decision Tree

The basic ideology for the DT model is the iterative dichotomizer (ID3) algorithm elaborated by Raza et al. [53]. However, this study used the following steps to develop the DT model: (1) input of T_{max} and T_{min} selected as predicted variables and (2) selection of the tree control parameters (i.e., node size, node split, and tree level). In this study, node size consisted of five rows, whereas the minimum split occurred at ten nodes, and the maximum size of the tree (level) was 10; (3) selection of the cross-validation method. In this study, V-fold cross-validation was employed as recommended by a previous study [53]; (4) an essential step is the selection of the pruning method; this study applied tree pruning control up to a minimum cross-validated error created in the earlier step; (5) finally, the target (ET_o) value was determined.

3.1.2. Tree Boost

A TB model contains several individual trees like SDTs in series. These smaller trees are interlinked with each other. Therefore, the connection is present between the adjacent trees. Thus, it is a chain-like structure with several trees. For the present study, the following steps were considered to develop the TB model: (i) input parameters of the climatic variables (T_{max} , T_{min}) assigned as predictor variables; (ii) commonly, misclassification is not crucial in ML, but its rate (%) has decided the predictive ability of ML model. In this way, the Huber loss function was applied to remove the misclassification error generated during the model development process; (iii) the surrogate splitter method was utilized to split the selected dataset; (iv) the minimum and the maximum number of trees in the TB model for creating a series to perform an ensemble process were 10 and 400, respectively, which were assigned after several attempts (trial and error method); (v) the dataset was segregated using the K-type classification, which divides the dataset into k and k-1 subsets. Each of the k subsets was tested, while the k-1 subset was deployed for training. The average value from each subset was collected and then the overall ET_o value as an output of the model was projected.

3.1.3. Decision Tree Forest

For the present study, the steps involved in the development of the TF model were as follows: (1) climatic variables (T_{max} , T_{min}) were selected as predictors/input; (2) 70% of the total dataset were labeled as sample data while 30% fell in the out-of-bag (OOB) data; (3) the sample data were split by the automatic randomization function used in the TF model; (4) the actual result of each tree was collected while residual variance during the process was collected separately; (5) the OOB data were analyzed and the resulting residual variance was separated; (6) the average values from all the resulting residual variance obtained from the sample and OOB were computed; (7) at last, the prediction was determined. The stepwise process for ML models based on a tree-like structure is presented in Figure 3. It can be seen that climatic variables (T_{max} , T_{min} , RH_{avg} , U, and N) are used as input to train tree-based ML models. In addition, a summary of the DT, TB, and TF models and their optimal parametric values are presented in Table 2.



Figure 3. Stepwise process for tree-based ML models on ET_o modeling.

Table 2. Outline of tree-based ML models.

NT N. 1.1.	Pasis Algorithm	Parametric Values for Model Development					
NIL Models	basic Aigorithm	Depth of Tree	Split Value	Pruned Size Node			
DT	Iterative Dichotomiser 3 (ID3)	09	26	10			
TB	Gradient Boosting Algorithm (GBA)	05	8.2	05			
TF	Random Forest Algorithm (RFA)	16	77.3	08			

3.2. Neural Network (NN)-Based ML Models

3.2.1. Multilayer Perceptron Neural Network (MLPNN)

The following steps were followed to develop an ET_o model using the MLPNN: (1) input variables of climate (T_{max} , T_{min}) were selected as predictors; (2) the activation functions of sigmoid and linear in the input-hidden and hidden-output layers were applied to process the dataset; (3) a back propagation procedure was adopted to adjust the weight connection between interconnected neurons; (4) the traditional conjugate gradient (TCG) and the scaled conjugate gradient (SCG) were employed as kernel functions; (5) the predicted ET_o was the model output.

3.2.2. Generalize Regression Neural Network (GRNN)

The GRNN was applied using the following steps [54]: (1) the reciprocal and Gaussian kernel functions were separately applied to investigate each performance; (2) minimizing the neurons in the network was considered the best way to optimize model evaluation, so the leave-one-out (LOO) method was used for this purpose; (3) the range of sigma from 0.0001 to 10 with search step of 20 was specified for each variable.

3.2.3. Cascade Correlation Neural Network (CCANN)

For the development of the CCANN model, the succeeding steps were as follows: (1) transform functions of the sigmoid (sig), Gaussian (GU), and a combination of sigmoid and Gaussian functions were used to process dataset information; (2) V-fold cross-validation method was employed to evaluate the network model size; (3) the network parameters of candidate neurons and epochs were 12 and 200 with a weight range of 1.00 after the hit and trail procedure.

3.2.4. Radial Basis Function Neural Network (RBFNN)

Shoaib et al. [55] elaborated the neural network of the RBFNN model in detail. The development of the RBFNN network for the present study was performed as follows: (1) the network parameters of maximum neurons, radius, and lambda were selected as 100, 400, and 10, respectively; (2) V-fold cross-validation was used to validate the dataset; (3) the tuning parameters of population size and maximum generation for the RBFNN network were 200 and 20, respectively; (4) the activation function of the radial basis function was employed in both hidden and output layers. The schematic structure of the applied NN-ML models is shown in Figure 4a,b, indicating the input–output relationship of temperature variables and ET_o for the current study. Details regarding the schematic structure of NN-ML models could be found in Raza et al. [44].





3.3. Multifunction (MF)-Based ML Models

3.3.1. Gene Expression Programming (GEP)

Shoaib et al. [56] comprehensively elaborated the theoretical background and architecture of GEP along with the stepwise procedure. Apart from theory, the following steps were chosen to develop the GEP model in the present study: (1) Root relative square error (RRSE) found the best

metric and chosen as the optimal fitness function; (2) climatic variables (T_{max} , T_{min}) were assigned as a set of terminals; (3) set of functions was chosen based on arithmetic (+, -, ×, ÷) and mathematical { $ln, e^x, \sqrt[2]{x}, x^2, x^3, x^4, Sinx, Cosx, Tanx, Sinhx, Coshx, Tanhx, ArcSx, ArcCx, ArcTx$ } operators; (4) two linking functions of addition (+) and multiplication (×) were selected; (5) finally, mathematical ET_o equations of various kinds were obtained by choosing the following general and evolution GEP parametric values: population size \rightarrow 50; No. of genes per chromosome \rightarrow 04; head length \rightarrow 8; inversion rate \rightarrow 0.1; transposition rate \rightarrow 0.1; recombination rates \rightarrow 0.1; one-point rate \rightarrow 0.3; two-point rate \rightarrow 0.3; mutation rate \rightarrow 0.044; IS transposition rate \rightarrow 0.1; RIS transposition rate \rightarrow 0.1.

3.3.2. Support Vector Machine (SVM)

The background knowledge of SVMs was described by Raza et al. [54]. Some steps were performed to develop the SVM model for ET_o , as follows: (1) two types of SVM models, namely, epsilon (\in) and neuron (Nu), were applied using the five climatic parameters (T_{max} , T_{min}); (2) various types of kernel function including linear, RBF sigmoid, and polynomial were tried; (3) model validation was determined by applying the V-fold cross-validation method; (4) pattern and grid search control with an interval of 10 was applied to determine accurate values for optimal parameters; (5) the parametric summary of the SVM model in term of ranges to predict ET_o is given as C (0.1–500); Nu (0.001–0.6); gamma (0.001–50); P (0.0001–100); coefficient (0–100); degree (3.00).

3.3.3. Global Method of Data Handling (GMDH)

Raza et al. [54] described the framework of the GMDH. The following steps were performed to determine ET_o using the GMDH model: (1) the maximum network layers of 20 with polynomial order 16 were assigned; (2) 20 neurons were fixed for each network layer; (3) V-fold cross-validation was used; (4) the previous layer in the GMDH network was connected to the initial output variables; (5) like the GEP model, various types of functions (linear, quadratic, cubic, product, ratio, asymptotic, logistic, Gaussian, logarithmic, exponential) with one and two variables were applied to generate corresponding ET_o equations. The topology of multifunction models for modeling the ET_o is shown in Figure 5.



Figure 5. Topology of multifunction (MF)-based ML models in ET_o modeling.

3.4. Proposed ML Models

Figure 6 depicts the flowchart of proposed ML models for ET_o estimation. It can be perceived that three tree-based, four NN-based, and three MF-based ML models were chosen to calculate ET_o using only temperature data (T_{min} , T_{max}) as input variables. The performance of each ML model was investigated by calculating different statistical evaluation indices.



Figure 6. Proposed ML flowchart for ET_o modeling.

3.5. Model Evaluation Indicators

The previous studies [40–50] applied several statistical indices to evaluate the performance of each ML model versus the conventional method (e.g., the FAO PM56 method). In the current study, the following statistical indices, root-mean-square error (*RMSE*), determination coefficient (R^2), mean absolute error (*MAE*), Nash–Sutcliffe efficiency (*NSE*), and correlation coefficient (r), were used to evaluate 10 ML models:

$$R^{2} = \left[\frac{\sum_{i=1}^{n} (ET_{obs} - \overline{ET}_{obs}) (ET_{est} - \overline{ET}_{est})}{\sqrt{\sum_{i=1}^{n} (ET_{obs} - \overline{ET}_{obs})^{2}} \sum_{i=1}^{n} (ET_{est} - \overline{ET}_{est})^{2}}\right]^{2}$$
(4)

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{\left(ET_{obs} - ET_{est}\right)^2}{n}}$$
(5)

$$r = \frac{\left[n \times \{\sum_{i=1}^{n} \{(ET_{obs} \times \overline{ET}_{est})\}\right] - \left[\{(\sum_{i=1}^{n} ET_{obs}) \times (\sum_{i=1}^{n} ET_{est})\}\right]}{\sqrt{\left[n \times \{\sum_{i=1}^{n} (ET_{obs})^{2}\right] - \left[\{(\sum_{i=1}^{n} ET_{obs})\}^{2}\right] \times \left[n \times \{\sum_{i=1}^{n} (ET_{est})\}^{2}\right] - \left[\{(\sum_{i=1}^{n} ET_{est})\}^{2}\right]}}$$
(6)

$$MAE = \sum_{i=1}^{N} \frac{|ET_{obs} - ET_{est}|}{n}$$
(7)

$$NSE = 1 - \frac{\sum_{i=1}^{n} (ET_{obs} - ET_{est})}{\sum_{i=1}^{n} (ET_{obs} - \overline{ET_{obs}})}$$
(8)

where ET_{obs} , ET_{est} , $\overline{ET_{obs}}$, $\overline{ET_{est}}$ are the observed value, the estimated value, the average observed value, the average estimated value of ET_o , respectively, and *n* represents the total number of observations.

4. Training Results of Developed ML Models

To accurately estimate the monthly ET_{o} , 10 different ML models were applied (MLPNN, GRNN, CCANN, DT, TF, GEP, GMDH, RBFNN, SVM, and TB). Since the data were collected from different locations and sites and had different characteristics, it was a challenge to create a reliable predictive model. The dataset was randomly grouped into phases called the training set and the testing set. The training set was used for the calibration process and model construction, while the testing set was used to examine each ML model's performance accuracy. Tree base-, NN-, and MF-based ML models were trained using only temperature variables (T_{max} and T_{min}), and ET_{o} estimated by the FAO PM56 method was used as the target variable. The ET_{o} estimated by each ML model was recorded and compared with PM ET_{o} . The obtained results as statistical indices (*RMSE*, *R*², *MAE*, *NSE*, and *r*) are explained below.

4.1. Tree-Based Models

Tree-based ML models of DT, TB, and TF were applied for ET_o estimation using climatic data from 15 stations. Climatic parameters of T_{min} and T_{max} were used as input to the selected models. The training results of these models are depicted in Table 3.

Model	Depth of the Tree	R ² (%)	r	<i>RMSE</i> (mm/Month)	MAE (mm/Month)	NSE (%)
DT	09	89.78	0.94	0.758	0.75	85.67
TB	05	96.87	0.98	0.41	0.42	95.34
TF	16	90.40	0.96	0.73	0.62	89.42

Table 3. Results obtained in the training process for the selected tree models.

A minimum depth of 5 was obtained for the TB model (Table 3). It is worth noting that the tree model with its minimum depth does not contain a complex structure as found in high-value cases. It is clear from Table 3 that TB performed best compared to the DT and TF models. R^2 and NSE for TB were 96.87% and 95.34%, respectively, while TF and DT had 90.40%, 89.78%, and 89.42% and 85.67%, respectively. In addition, the correlation coefficient (*r*) of TB was higher, i.e., 0.98, compared to the DT and TF models. On the other hand, *RMSE* and *MAE* were lower for the TB model, indicating the best performance among other tree base-chosen models.

4.2. Neural Network (NN)-Based ML Models

Table 4 indicates the applied neural network-based models of MLPNN, GRNN, CCANN, and RBFNN for ET_0 estimation using various kernel functions. For the MLPNN model, the kernel function of the traditional conjugate gradient (TCG) with optimal structure 2-18-1 outperformed the scale conjugate gradient (SCG). R^2 and r for the TCG kernel function were the highest values, while *RMSE* and *MAE* were lower as compared to SCG. Likewise, in the GRNN model, Gaussian (Gu) and reciprocal (Res) functions were applied to determine ET_0 by employing climatic variables. It can be inferred from Table 4 that the GRNN with the optimal structure of 2-5-1 at the Res kernel function was found best in performance and showed promising results. The correlation (r) and determination

coefficient (R^2) were 0.99 and 99.99%, correspondingly, for Res, which is higher than the Gu function. Conversely, the error at the Gu function recorded higher was as follows: RMSE = 0.29 mm/month, MAE = 0.18 mm/month. However, NSE during training was recorded as 97.43% and 98.67% for the Gu and Res functions, respectively. This indicates that the performance of the GRNN model was reasonable at both kernel functions and can be used to estimate ET_0 .

Model	Kernel Function	NN Structure	R ² (%)	r	<i>RMSE</i> (mm/Month)	MAE (mm/Month)	NSE (%)
MLPNN	SCG	2-6-1	85.78	0.92	0.89	0.66	84.72
	TCG	2-18-1	88.02	0.93	0.82	0.60	87.17
GRNN	Gu	2-7-1	98.41	0.99	0.29	0.18	97.43
	Res	2-5-1	99.99	0.99	0.01	0.01	98.67
CCANN	Sig	2-8-1	98.92	0.99	0.24	0.18	97.88
	Gu	2-12-1	97.59	0.98	0.36	0.30	96.48
	S&G	2-16-1	96.23	0.98	0.46	0.36	95.24
RBFNN	RBF	2-36-1	96.41	0.98	0.44	0.35	95.32

Table 4. Summary outlier of the selected neural network-based ML models.

Likewise, the results of the CCANN model at three kernel functions of Gu, Sig, and a combination of sigmoid and Gaussian functions are summarized in Table 4. The results indicated that the minimum number of hidden neurons with an optimal structure of 2-8-1 by using the sigmoid function in the CCANN model outperformed the Gu and S&G kernel functions. For the Sig kernel function, model performance indices of the CCANN were calculated as follows: $R^2 = 98.92\%$; r = 0.24; RMSE = 0.24 mm/month; MAE = 0.18 mm/month; NSE = 97.88%. On the other hand, the CCANN with the Gaussian function generated higher R^2 (97.59%) and lower RMSE (0.36 mm/month) compared to the S&G kernel function with an optimal structure of 4-12-1. In addition, the RBFNN network with the radial basis function (RBF) was applied to estimate ET_o. The output results of the RBFNN can be seen in Table 4. It can be observed that the RBFNN generated reliable estimates at the optimal structure of 2-36-1 as follows: $R^2 = 96.41\%$; r = 0.98; RMSE = 0.44 mm/month; MAE = 0.36 mm/month; and NSE = 95.32%.

Overall, it can be concluded that the MLPNN at the TCG, the GRNN at Res, the CCANN at Sig, and the RBFNN at the RBF kernel function showed the best performance among others for estimating ET_0 . Our results aligned with the finding of Martí et al. [57] who stated that a model with minimum neurons was selected and considered best for ET_0 estimation. In this study, the GRNN at Res outperformed others, with minimum neurons among all the neural network-based models, as verified in Table 4.

4.3. Multifunction (MF)-Based ML Models

The results of the SVM for the training period are depicted in Table 5. It can be seen that \notin -SVM and Nu-SVM were applied at different kernel functions. Both types of the SVM showed outperformed the results at the RBF kernel function compared to others. The R^2 and r for both SVMs at RBF were to be the highest among others. For \notin -SVM, R^2 was observed at 99.77% while r was estimated as 1.00. Similarly, for Nu-SVM, R^2 and r were noted as 99.66% and 1.00, respectively. As shown in Table 5, higher values of *RMSE* and *MAE* were generated when applying the SVM model at linear, sigmoid, and polynomial kernel functions. These are listed as follows: 2.08 mm/month and 1.78 mm/month of *RMSE* and *MAE* for the linear kernel integrated with both SVM models; *RMSE* = 2.10 mm/month, *MAE* = 1.78 mm/month and *RMSE* = 2.08 mm/month, *MAE* = 1.78 mm/month for the sigmoid kernel with \notin -SVM and Nu-SVM; *RMSE* = 2.16 mm/month, *MAE* = 1.89 mm/month; and *RMSE* = 2.23 mm/month, *MAE* = 1.78 mm/month for the sigmoid kernel function with \notin -SVM and Nu-SVM; *RMSE* = 2.16 mm/month for the polynomial kernel function with \notin -SVM and Nu-SVM; *RMSE* = 2.16 mm/month and *RMSE* = 1.78 mm/month for the sigmoid kernel with \notin -SVM and Nu-SVM; *RMSE* = 2.16 mm/month, *MAE* = 1.89 mm/month; and *RMSE* = 2.23 mm/month, *MAE* = 1.78 mm/month for the sigmoid kernel with \notin -SVM and Nu-SVM models, accordingly.

SVM Model	Kernel Function	R ² (%)	R	<i>RMSE</i> (mm/Month)	MAE (mm/Month)	NSE (%)
	Linear	73.98	0.86	2.08	1.78	72.49
€-SVM	RBF	99.77	1.00	0.11	0.10	98.56
	Polynomial	65.46	0.80	2.16	1.89	64.29
	Sigmoid	69.67	0.83	2.10	1.78	68.55
	Linear	69.26	0.83	2.08	1.78	67.32
	RBF	99.66	1.00	0.14	0.13	98.49
Nu-SVM	Polynomial	60.32	0.77	2.23	1.78	59.79
	Sigmoid	67.25	0.82	2.08	1.78	66.98

Table 5. Results of the applied SVM model at each kernel function (training period).

The results obtained from the GEP are presented in Table 6. The present study subtilized the addition (Add.) and multiplication (Mul.) linking functions with various mathematical and arithmetic operators to calculate the ET_o equation. All the operators were labeled from F1 to F20 (Appendix A.2). It can be inferred from Appendix A.2 that the GEP model outperformed and showed promising results with the Add. linking function and F8 operators {+, -, ×, ÷, *Sinhx*, *Coshx*, *Tanhx*}. The *RMSE* obtained at this combination was 1.44 mm/month, the recorded minimum. The GEP generated good results for ET_o estimation on the addition (Add) and multiplication (Mul) linking functions at the F8 and F11 operators, respectively (Table 6). These were highly ranked among others as the lowest values (1.44 mm/month and 1.47 mm/month) of *RMSE* obtained. The corresponding R², r, MAE, and NSE results were also presented in Table 6.

Table 6. Results of the GEP for the best operators under the Add. and Mul. linking functions (Appendix A.2).

No.	Operators	Linking Function	R ² (%)	r	<i>RMSE</i> (mm/Month)	MAE (mm/Month)	NSE (%)
F8	$ \{+, -, \times, \div, Sinhx, Coshx, Tanhx\} \\ \{+, -, \times, \div\} $	Add.	82.98	0.91	1.44	1.55	81.49
F11		Mul.	80.77	0.89	1.47	1.62	79.56

Like the GEP model, the GMDH applied for ET_o estimation used multiple types of functions (i.e., linear, quadratic, product, exponential, etc.). The mathematical description of each function in terms of the equation could be determined from Appendix A.3. Additionally, *RMSE* was selected as an evaluation parameter for each function. The *RMSE* for the GMDH model was found in the range of 0.88 to 2.23 mm/month. The evaluation indices (R^2 , *r*, *RMSE*, *MAE*, and *NSE*) were recorded at a quadratic function with two variables of the GMDH model (Table 7).

Table 7. Results of the GMDH for ET_o estimation at the best function (Appendix A.3).

Function	Equation	R ² (%)	r	<i>RMSE</i> (mm/Month)	MAE (mm/Month)	NSE (%)
Quadratic 2 variables	$ET_o = y$	91.06	0.95	0.83	0.85	90.88
Note: $V = n + n + n + n + n + n^2$	$ n x n x^2 $	n x x (co	officiante	n1 n5 given in A	ppopdix (12)	

Note: $Y = p_1 + p_2 x_1 + p_3 x_1^2 + p_4 x_2 + p_5 x_2^2 + p_6 x_1 x_2$ (coefficients p1–p5 given in Appendix A.3).

5. Evaluation of the Proposed ML Models against the FAO PM56 Method

This study evaluated the trained ML models using climatic parameters of T_{max} and T_{min} only to estimate ET_o at seven climatic stations in Pakistan (Gilgit, Islamabad, Jacobabad, Karachi, Lyallpur, Multan, and Skardu). The climatic and geographic details corresponding to each selected climatic station are mentioned in Appendix A.1. ET_o estimated with the MLPNN, GRNN, CCANN, DT, TF, GEP, GMDH, RBFNN, SVM, and TB were labeled as MLPNN ET_o , GRNN ET_o , CCANN ET_o , DT ET_o , TF ET_o , GEP ET_o , GMDH ET_o , RBFNN ET_o , SVM ET_o , and TB ET_o , respectively. The flowchart of proposed

ML models for ET_o estimation can be seen in Figure 6. The ET_o obtained from the applied ML models were compared with PM ET_o values to investigate their performance. For this purpose, evaluation indices (r, R^2 , RMSE, MAE, and NSE) were calculated and presented in Figure 7. However, scatterplots between the actual (PM ET_o) and proposed ML models (predicted) are given in Figure 8. It can be perceived from Figures 7 and 8 that the proposed ML models generated the best results for ET_o estimation. However, GRNN ET_o, and TB ET_o were found to be very close to PM ET_o. In addition, r, R^2 , and NSE values could be found at nearly 1.00. Alternatively, RMSE and MAE were found lower than with the other ML models. For the GRNN model, r = 0.97, $R^2 = 0.96$, NSE = 0.99, MAE = 0.36 and RMSE = 0.46, while r = 0.99, $R^2 = 1.00$, NSE = 1.00, MAE = 0.26 and RMSE = 0.37 for the TB model.



Figure 7. Results of evaluation indices (r, R², NSE, RMSE and MAE) for the applied ML models.

Figure 9 showed that ET_o estimated by ML models outperformed and followed PM56 ET_o . Overestimation of ET_o values was found using the GEP and RBFNN models due to low correlation and high variance. On the other hand, the DT model showed a high value of RMSE but generated good ET_o values as it has high efficiency. Therefore, ET_o estimated using DT appeared closer to the FAO PM56 method than with the GEP and the RBF, as observed in Figure 9. The overall results indicated that TB ET_o followed well with PM ET_o and was capable of capturing monthly ET_o variation. In comparison of the tree-based, neural network-based, and multifunction-based ML models, it was found that TB (among the tree-based), GRNN (among the neural network-based), and SVM (among the multifunction-based) outperformed all the applied ML models.

The results of this study aligned with the results of Mohammadrezapour et al. [40] who stated that tree-based ML models outperform others in an arid region (Sistan and Baluchestan Province) of Iran using only temperature climatic variables. Likewise, Sanikhani et al. [39] calculated reliable ET_o by applying the GEP and GRNN using temperature data only in the Mediterranean region (hot and dry during the summer season and rainy during the winter season) of Turkey. Similarly, Granata [46] proved the performance of ML models against the FAO PM56 method using limited data of temperature variables at a central Florida site located in a humid climate. Thus, it can be concluded that ML models showed good performance in various climatic conditions using only time series temperature data, which also supports our study results.



Figure 8. Scatterplot of ML models during the evaluation phase.



Figure 9. ET_o comparison of the applied ML models with the FAO PM56 method.

A comparison of ET_o obtained from the FAO PM56 method and ML models using temperature data only for the stations mentioned above is presented in Table 8 to display the ET_o results. It can be concluded that the selected ML models outperformed others using temperature data only and showed good results under different climatic conditions. The overall results indicate that TB (among the tree-based), GRNN (among the neural network-based), and SVM (among the multifunction-based) performed best at the tested climatic stations and yielded approximate results for ET_o estimation compared to PM ET_o . However, TB ET_o trailed well with PM ET_o , as observed in Table 8.

Model	Gilgit	Islamabad	Jacobabad	Karachi	Lyallpur	Multan	Skardu
PM ET _o	2.7	5.1	6.43	4.65	4.38	5.38	2.69
MLPNN ET _o	2.5	5.0	6.53	4.68	4.32	5.19	1.86
GRNN ET _o	2.7	5.1	6.45	4.64	4.38	5.37	2.69
CCANN ET _o	2.8	5.3	6.51	4.71	4.43	4.92	2.37
RBFNN ETo	2.5	5.0	6.51	4.76	4.21	5.40	1.87
SDT ET _o	3.2	5.0	6.40	4.94	4.90	5.94	2.44
DTF ET	3.2	5.0	6.67	4.90	4.91	6.10	2.61
TB ET _o	2.7	5.1	6.42	4.64	4.38	5.37	2.69
GEP ET _o	2.6	5.2	6.35	6.07	5.19	5.70	2.33
GMDH ET _o	3.0	5.2	6.59	4.64	4.57	5.05	2.11
SVM ET _o	2.9	5.2	6.40	4.69	4.36	5.40	2.72

Table 8. Comparison of PM ET_o with ML ET_o.

The heatmap diagram, as presented in Figure 10 provided significant information about the best modeling performance based on five statistical indices (r, R², NSE, RMSE, and MAE) and also supported our above-mentioned results, which confirmed the TB ML model as an outstanding model to estimate ET_{o} using only temperature data. Although GEP generated acceptable model efficacy (precision) during the training phase (*NSE* = 81.49%), it exhibited the worst accuracies (*NSE* = 28%) as compared to other ML models in the testing set. Additionally, the other statistical indices, such as RMSE and MAE, also gave further information about the weaknesses of GEP model. It is true that the GEP ML model suffers from an overfitting issue. On the other hand, the RBFNN and MLP showed much better performance than GEP. Finally, the TB and SVM ML models achieved high precision in ET_{o} estimation with the FAO PM56 method. Moreover, the TB ML model is considered the best predictive model and recorded the highest $R^2 = 1$; *NSE* = 1; r = 0.99 and lowest *MAE* = 0.26 mm/month, *RMSE* = 0.37 mm/month, respectively.



Figure 10. Selection of Best ML model through a heatmap diagram.

Considering the ML models, it can be noted that this approach often requires a relatively lower number of input parameters compared to FAO PM56 method. The logical explanation of this phenomenon is that the data size (training and testing) includes total 22 locations of different characteristics. Besides, geographical factors are very important when developing robust ML models based on a dataset collected from several locations. For further assessment, it is vital to examine the ability of the proposed model (TB) against several ML models developed in the past to estimate ET_0 globally. Thus, the findings obtained during the test using the TB model are validated against various ML models carried out in the literature. For example, Raza et al. [53] compared tree-based ML model in ET_0 estimation by considering 11 stations from different parts of the world and found in the testing phase that TB performed better than DT and TF ML models based on R² and RMSE indices. Likewise, Raza et al. [54] showed the supremacy of MLP over several ML models in ET_0 estimation by assessing various performing indices at various climatic locations.

*ET*_o Interpolation Maps Based on the Best ML Model

ArcMap GIS 10.1 software was used for mapping ET_o of Pakistan. For this purpose, an interpolation technique was used to determine ET_0 values using point data of ET_0 . In addition, interpolation is a technique to find unknown values using available sample dataset values. It can be used to predict an unknown value for any geographic dataset. Different tools existed like Inverse Distance Weighted (IDW), Kriging and Natural Neighbor for interpolation of point data. This study applied the IDW tool to prepare the surface raster map by employing TB ET_0 point data (Appendix A.4) and geographic data. This IDW-interpolation method is a straightforward and unique method based on an average value of the sample data algorithm. Figure 11 shows interpolated monthly ET_o maps of Jan–Dec, while Figure 12 showed Pakistan's annual ET_0 map, respectively. These ET_0 maps are based on temporal and geographic data and present realistic scenarios, which are mainly helpful to determine crop water requirements and obtain good crop yield with saving water. According to Figure 11, ETo increased from Jan to Jul while started to drop from Aug to Dec. The lowest ET_0 recorded in Jan (0.63) which gradually increased 4.49 up to Jul month and same trend could be observed (4.19 to 11.83) in the highest ET_0 from Jan–Jul. However, Aug-Dec months showed decreased in ET_o on both low and high scale. From Figure 2, it can be perceived that temperatures (T_{min}, T_{max}) showed same inclining (Jan-Jul) and declining (Aug-Dec) trends which showed significant relationship with ET_0 . From Figure 12, ET_0 increases from the southern to the northern part of Pakistan. The peak of ET_o was found in the May–Jul months. The southeastern region looked more affected than other regions and recorded the highest ET_o. From Aug–Dec months, the overall effects of ET_o were reduced in the southern and southeastern parts of Pakistan.



Figure 11. Monthly (Jan–Dec) ET_o map of Pakistan based on geographic data.



Figure 12. Annual ET_o map of Pakistan based on geographic data.

6. Study Discussion and Comparison with ML Studies

Certainly, ET_o is affected by near-surface and ground conditions, but temperature data primarily influence it. This study found a high correlation between ET_o and temperature data because it also incorporates effects of other local climatic variables (i.e., RH, U, N, solar radiation (R_s), etc.). ET_o also changed with latitude/locations and showed location has strong impact on the spatiotemporal variations of ETo. Wang et al. [58] determined the effect of altitude and latitude on surface air temperature across the Qinghai-Tibet Plateau and found that temperature variations depend on not only altitude but also on latitude. There is a gradual decrease in temperature with the increasing altitude and latitude. In Turkey, Turgut and Usanmaz [59] investigated variations in wind speed, wind direction, and humidity, depending on altitude. They found that average wind direction changes from 169° at the lowest altitude to 260° at the highest altitude. Also, average wind speed increases by 2.3 m/s, while average relative humidity decreases by 4%, for an increase of one kilometer in altitude [59]. One of the most important reasons for advocating a simpler method than FAO PM56 is the substantial likelihood of inaccuracy in weather data measurement and collection, especially in developing countries (e.g., Pakistan) and meteorological stations managed by non-experts. In these situations, data accuracy and quality parameters may not be reliable [60]. The discussion presented below also support our study's results based on statistical indices.

The capacity of different ML to model monthly and daily mean ET_o using temperature data alone from local or cross stations was investigated by Wu et al. [41]. For the local application, tree-based ML models had greater estimation accuracy based on statistical indices ($R^2 = 0.962$ and RMSE = 0.263 mm/day) than the other models. SVM was appropriate model where temperature data is not available at the site, and data from nearby stations can be utilized. If meteorological data is missing, the tree-based ML can provide the best solution for ET_o modeling. Likewise, Mohammadrezapour et al. [40] applied SVM, GEP, and adoptive neuro-fuzzy interference system (ANFIS) using different input climatic combinations for ET_o modeling. The results showed that SVM performed best with $R^2 = 0.998$ and

RMSE = 0.434 mm/month among the selected ML models. Saggi and Jain [42] examined four ML models, namely, deep learning (DL), generalized linear model, gradient boosting machine, and TF for daily ET_o modeling in India. The results showed superior performance of the DL model with the highest NSE of 0.98 and lowest RMSE of 0.19 compared to other models. Shiri et al. [43] compared GEP with locally and externally calibrated Priestley Taylor (PT) model on ET_o estimation. They found that GEP outperformed (with indices of *RMSE* = 0.462 mm/day; *MAE* = 0.216 mm/day) and provided the best solution for ET_o modeling alternative to the FAO PM56 method using two meteorological inputs in humid and arid stations of Iran. Tikhamarine et al. [44] applied five integrated algorithms of ML neural network for modeling monthly ET_o at two stations located in India and Algeria. The obtained ML results compared with Valiantzas model based on an empirical equation. It was found that ML with grey wolf optimization algorithms (ML GWO) performed best in comparison to the empirical equation as indices calculated at both stations as *NSE* = 0.99 and *RMSE* = 0.05–0.08 mm/month.

Ferreira et al. [45] stated that most empirical equations commonly reported are sitespecific or less extensive climatic conditions, limiting their use globally. Therefore, the authors applied several ML models for ET_0 modeling in the entire region of Brazil using fewer climatic data. The authors recommended that ML is the best choice in case of missing meteorological data. In addition, the study concluded that the absence of RH data in analysis decreases RMSE up to 24%. Granata [46] determined the best ET_0 ML model among SVM and tree base (DT, TB, TF) by using climatic data of the humid region in Florida. Keshtegar et al. [47] developed a polynomial chaos expansion (PCE) ML model for ET_o modeling using limited climatic inputs at two stations in Turkey. The results demonstrated that the PCE ML model outperformed other alternative approaches and generated the highest NSE of 0.999, lowest RMSE of 0.045 mm, and agreement index of 0.999. Nourani et al. [48] compared ML (neural network and multifunction) and empirical models for ET_o modeling in various climatic regions globally (e.g., Turkey, Iraq, Cyprus, Iran, and Libya). The results of ML models showed supremacy over empirical models. Additionally, two ensemble models based on ML and empirical equations were also developed to improve ET_o results and compared with single ML and empirical equations. The study recommends using the ML model in case of missing climatic data for ET_o modeling. Likewise, Shiri [49] investigated a multifunction ML (GEP) model compared to six empirical equations for daily ET_0 modeling using island climatic data in Iran. The results of GEP suppressed over the selected empirical models and externally calibrated GEP model performed best at the test island because there was no need to train GEP model with local climatic inputs, hence generated reliable ET_o.

Kisi et al. [61] investigated the ability of four ML models for monthly ET_o modeling in Iran. GEP ML model did not perform well at several stations and provided the worst estimate in this study, which contradicts our study results. This study showed good performance of ML GEP model using input data, which may be due to the use of optimal parametric values and the best linking function during training of the GEP model. Similarly, neural network and ANFIS models outperformed and provided a better estimate for ET_o, which agrees with our study. However, applications of ML over empirical and local calibrated models are recommended to use in case of missing climatic data. In addition, over and under estimation in ET_o values using ML models are depends on proper calibration in the training phase. Higher training data leads to undereste ET_0 while ET_0 overestimation was observed in the case of minimum training data. As the current study used the least amount of training points due to the limitation of the monthly-scale data, under-estimation in ET_o was also observed using GEP and DT ML models against the FAO PM56 method. Marti et al. [62] examined the ML model's performance in case of missing climatic data at various climatic stations in India. The application of ML models with minimal data needs proper training, and the use of ET_o models based on ML can be applied to other climatic conditions for its verification and validation. Therefore, this study investigated ML models in various climatic conditions to verify the performance of developed ML ET₀ models. Table 9 shows data requirements for FAO PM56 and ML models for ET_o estimation. It can be seen in Table 9 that FAO PM56 depends upon several parameters, including other factors, which are not easily accessible, especially in developing countries. Alternatively, ML models require fewer parameters (only temperature data) that give the best ET_o value compared to the FAO PM56 method. The symbol " $\sqrt{"}$ " in Table 9 represents the parameters needed for ET_o estimation while "-" indicates not used in the corresponding procedure.

Table 9. Data requirements for FAO PM56 and ML models in ET_o estimation.

Input Data Parameter	T _{min}	T _{max}	RH	U	N	Rn	Aerodynamic Factors (R _n , es, ea, e _{min} , emax, Δ , Z, and γ)	Adopted Methodology	Target Result
Climatic and aerodynamic	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	FAO PM56	PM ET _o
Temperature	\checkmark	\checkmark	-	-	-	-	-	ML models	ML ET _o

Significant attempts were made to increasingly constitute powerful models to accurately estimate ET_0 based on a few parameters that can easily be measured. Zhu et al. [63] tried to build a ML model based on temperature data only and compared its ET_0 predictive results with several empirical equations. The study examined that the developed ML model performed best by producing the least error in the testing phase and showed a high correlation ($R^2 = 89\%$) value between actual and predicted ET_0 . Overall, as shown above, the approach of past research has been used to estimate ET_0 over a number of sites. Even though researchers produced powerful ML models with excellent accuracy, the methodologies discussed could not be used to form a generic model instead of estimating each case study independently. However, the suggested ML model accurately estimated the ET_0 using data from dozens of sites in our investigation. Aside from the training procedure of 15 stations in a single model, another intriguing characteristic is that the suggested model's efficacy and prediction accuracy were high (NSE = 1; $R^2 = 1$).

7. Conclusions

This study examined tree-based (DT, TB, TF), neural network-based (MLP, GRNN, CCANN, and RBFNN), and multifunction-based (GMDH, SVM, and GEP) machine learning (ML) models in the estimation of ET_o using temperature data of different climatic stations located in Pakistan. Based on the model evaluation statistics, the performance of each selected model was compared with standard ET_o estimated by the FAO PM56 method. The highest-ranked among the tree-based ML models was TB, which was found to outperform ten ML models. When temperature (minimum and maximum) data were used as an input to the ML TB model, the results were as follows: r = 0.99, $R^2 = 1.00$, NSE = 1.00, MAE = 0.26, and RMSE = 0.37. On the other hand, the GRNN at the joint kernel function performed best among the neural network-based ML models with the optimal structure of 2-5-1 and best-performing indices: r = 0.97, NSE = 99%, $R^2 = 95.00\%$, RMSE = 0.61 mm/month, and MAE = 0.50 mm/month. Various types of kernel function interlinked with \notin -SVM and Nu-SVM were tried to investigate accuracy of the results. The ϵ -SVM with the radial basis function (RBF) was found to be the best among the multifunction-based models with r = 0.99, NSE = 88%, $R^2 = 98.00\%$, RMSE = 0.37 mm/month, and MAE = 0.29 mm/month. In addition, the performance of the GMDH was found to be superior to the GEP by calculating the least variance and high correlation. Except for the GEP model, all the selected ML models had high efficiency, and NSE was mostly found above 90%. TB was found largely the best in the estimation of ET_o and close to PM ET_o using only temperature data. Thus, it can be concluded that the estimation of ET_0 could be efficaciously determined by using proposed ML models with available temperature data. The monthly and annually varying ET_o maps based on the ML TB model indicated that more attention is needed towards southern and southeastern parts of Pakistan for precise planning and management of water use.

Limitations, Suggested Improvements and Future Directions

The use of proposed ET_{o} ML models is limited to the study region using only temperature data. So, it is necessary to create similar or novel ML models using fewer or similar meteorological inputs to investigate the performance of the developed ML models in other regions. Moreover, the proposed ET_{o} model of the current study also requires the selection of proper training/calibration before applying it to other regions. In addition, ML models do not have physical mechanisms. They are called black-box techniques. In these techniques, a user only knows the input and the expected output of the model but is unaware of how the program achieves those results. Therefore, it is challenging to create an accurate ML model without knowledge of functional specifications. Moreover, over- and underfitting problems could also be found during the training/calibration process of an ML model due to the random division of the dataset.

The future developments of this study will concern the creation of more ET_o models based on hybrid data intelligence (HDI) techniques and extreme learning machine (ELM) that contemplate different climatic conditions and subsequently consider the effects of climate change if present. Efforts on gap infilling techniques are mandatory, as effective planning, management, and control of water resource systems require considerable and reliable data on ET_o modeling. Particular attention should be given to regional models due to their crucial role in places with missing data to develop local models. Important information on various water resources issues, such as examining water use by different regions, water rights, water allocation, water consumption, and planning and management of ground and surface water resources, can be managed by ET_o maps. The ET_o maps in different months revealed that the highest amounts of ET_o occurred in the southern parts of Pakistan. More attention should be focused on water resource planning of these parts.

Author Contributions: Conceptualization, J.W., A.R. and Y.H.; Data curation, A.R. and K.S.; Formal analysis, A.R. and A.E.; Funding acquisition, J.W., Y.H. and P.L.; Investigation, A.R., K.S. and A.E.; Methodology, J.W., A.R. and A.E.; Project administration, Y.H.; Resources, Y.H.; Software, A.R., M.S. and K.S.; Supervision, Y.H.; Validation, J.W., N.A.B. and P.L.; Visualization, J.W., N.A.B. and P.L.; Writing—original draft preparation, J.W.; Writing—review and editing, A.R., Y.H., M.S. and R.L.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key R&D program of Zhenjiang (NY2018007), China, Jiangsu Postdoctoral Science Foundations (2016M600376 and 1601032C), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD-2018-87).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on reasonable request.

Acknowledgments: We are heartedly thankful to the anonymous reviewers for their helpful comments to improve the article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Monthly Averages of Climatic Variables and PM ET_o

Sr. No.	Climatic Stations	Lon DD	Lat DD	Alt m	$T_{min} ^{\circ}C$	T_{max} °C	RH _{avg} %	U km/Day	N Hours	PM ET _o mm
1	Multan	71.43	30.2	123	18.6	33.3	57	184	9.3	5.38
2	Chaman	66.45	30.93	1313	12.7	25.7	36	196	7.9	4.89
3	Quetta	67	30.16	1672	6.2	24.1	54	313	10.6	5.14
4	Lyallpur	73.01	31.36	184	17.4	31.6	53	130	7	4.38
5	Zhob	69.46	31.35	1407	12	26.5	45	89	7.6	3.7
6	Sargodha	72.66	32.05	188	16.7	30.9	57	222	7.6	5.03
7	Chilas	74.11	32.41	1250	14.3	26.7	33	56	5.9	3.29

Sr. No.	Climatic Stations	Lon DD	Lat DD	Alt m	T _{min} °C	T_{max} °C	RH _{avg} %	U km/Day	N Hours	PM ET _o mm
8	Parachinar	70.08	33.86	1726	8.8	21.3	47	101	7.2	3.23
9	Islamabad	73.1	33.61	508	14.4	28.4	55	269	7.5	5.06
10	Peshawar	71.58	34.01	360	15.7	29.3	50	156	8.1	4.51
11	Karachi	66.98	24.8	04	20.3	31.6	72	338	7.4	4.64
12	Karachi	67.13	24.9	22	20.3	31.5	62	482	8	5.97
13	Gilgat	74.33	35.93	1454	11.2	22.4	46	37	5.9	2.71
14	Astore	74.9	35.36	2168	4.3	15.5	46	58	4.9	2.5
15	Sakardu	75.61	35.3	2181	4.6	17.1	56	62	6.1	2.69
16	Hyderabad	68.41	25.38	28	21.1	35.1	48	345	8.3	7.05
17	Gupis	73.4	36.16	2156	7.4	18.9	36	62	6	2.85
18	Nawabshah	68.36	26.25	38	18.3	34.8	49	303	7.9	6.69
19	Nokkndi	62.75	28.81	683	16.7	32.2	35	158	7.7	5.28
20	Dal Bandin	64.4	28.88	850	13	30.3	37	158	7.8	5.03
21	Jacobabad	68.46	28.3	56	20.6	34.4	47	266	7.8	6.43
22	Kalat	66.58	29.03	2017	3.7	22.3	43	183	7.9	4.17

Appendix A.2. Linking Functions for the GEP Machine Learning Model

No.	Operators	Linking Function	RMSE (mm/Month)
F1	$\{+,-,\times,\div\}$	Add.	1.94
F2	$\{+, -, \times, \div, \sqrt[2]{x}, x^2, x^3\}$	Add.	2.09
F3	$\{+, -, \times, \div, \sqrt[2]{x}, x^2, x^3, x^4\}$	Add.	2.11
F4	$\{+,-,\times,\div,ln,e^x\}$	Add.	2.03
F5	$\{+,-,\times,\div,\ln,e^x,\sqrt[2]{x},x^2,x^3\}$	Add.	1.78
F6	$\{+, -, \times, \div, ln, e^x, \sqrt[2]{x}, x^2, x^3, x^4\}$	Add.	1.63
F7	$\{+, -, \times, \div, Sinx, Cosx, Tanx\}$	Add.	1.55
F8	$\{+, -, \times, \div, Sinhx, Coshx, Tanhx\}$	Add.	1.44
F9	$\{+, -, \times, \div, ArcSx, ArcCx, ArcTx\}$	Add.	1.69
F10	$\{+, -, \times, \div, Sinx, Coshx, ArcTx\}$	Add.	1.46
F11	$\{+,-,\times,\div\}$	Mul.	1.47
F12	$\{+, -, \times, \div, \sqrt[2]{x}, x^2, x^3\}$	Mul.	2.13
F13	$\{+, -, \times, \div, \sqrt[2]{x}, x^2, x^3, x^4\}$	Mul.	1.97
F14	$\{+,-,\times,\div,ln,e^x\}$	Mul.	2.18
F15	$\{+,-,\times,\div,\ln,e^x,\sqrt[2]{x},x^2,x^3\}$	Mul.	2.13
F16	$\{+, -, \times, \div, ln, e^x, \sqrt[2]{x}, x^2, x^3, x^4\}$	Mul.	2.13
F17	$\{+, -, \times, \div, Sinx, Cosx, Tanx\}$	Mul.	1.51
F18	$\{+, -, \times, \div, Sinhx, Coshx, Tanhx\}$	Mul.	2.08
F19	$\{+, -, \times, \div, ArcSx, ArcCx, ArcTx\}$	Mul.	2.37
F20	$\{+, -, \times, \div, Sinx, Coshx, ArcTx\}$	Mul.	1.51

Appendix A.3. Linking Functions for the GMDH Machine Learning Model

_

Functions	Equations	RMSE (mm/Month)
Linear 1 variable	$y = p_1 + p_2 x_1$	2.11
Linear 2 variables	$y = p_1 + p_2 x_1 + p_3 x_2$	2.10
Linear 3 variables	$y = p_1 + p_2 x_1 + p_3 x_2 + p_4 x_3$	2.07
Quadratic 1 variable	$y = p_1 + p_2 x_1^2$	1.48
Quadratic 2 variables	$y = p_1 + p_2 x_1 + p_3 x_1^2 + p_4 x_2 + p_5 x_2^2 + p_1 p_3 x_2$	0.83
Cubic 1 variable	$y = p_1 + p_2 x_1 + p_3 x_1^2 + p_4 x_1^3$	1.55
Ratio 2 variables	$y = p_1 + p_2 \frac{x_1}{x_2}$	1.86
Asymptotic 1 variable	$y = \frac{p_1 + p_2}{x_1 + p_3}$	2.20
Gaussian 1 variable	$y = \frac{p_1 + p_2 exp(-(x_1 - p_3)^2)}{p_4}$	1.48
Logistic 1 variable	$y = \frac{p_1^{+} + p_2}{(1 + exp(-p_3(x_1 - p_4)))}$	2.10
Exponential 1 variable	$y = p_1 + p_2 exp(p_3(x_1 + p_4))$	2.14
Product 2 variables	$y = p_1 + p_2 x_1 x_2$	2.11
Log 1 variable	$y = p_1 + p_2 log((x_1 + p_3))$	2.23

Note: P1 = T_{max} ; P2 = T_{min} ; P3 = RH_{avg}; P4 = N; P5 = U and $x_1 = 0.876$; $x_2 = 0.7654$; $x_3 = 0.5576$.

Long	Lat	Alt	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
74.9	35.36	2168	0.69	0.88	1.46	2.53	3.47	4.45	4.4	4.18	3.43	2.32	1.42	0.72
66.45	30.93	1313	1.81	2.39	3.59	5.04	6.82	7.99	7.95	7.23	6.15	4.67	3.01	1.99
74.11	32.41	1250	0.91	1.4	2.39	3.52	4.31	5.4	6	5.62	4.53	2.87	1.53	1.01
64.4	28.88	850	1.88	2.55	4.07	5.58	7.06	8.11	8.3	7.74	5.93	4.21	2.85	2.06
74.33	35.93	1454	0.74	1.2	2.06	2.99	3.95	4.78	4.8	4.35	3.45	2.21	1.21	0.78
73.4	36.16	2156	0.7	1.05	1.99	3.23	4.14	5.24	5.07	4.67	3.76	2.37	1.25	0.75
68.41	25.38	28	3.62	4.44	6.41	8.43	11.19	11.52	11.83	8.16	8.06	6.57	4.43	3.58
73.1	33.61	508	1.85	3.02	3.86	6.32	8.71	9.99	6.78	5.47	5.19	4.31	3.15	2.07
68.46	28.3	56	2.89	3.95	5.89	8.1	10.34	10.76	8.83	7.58	6.8	5.63	3.73	2.64
66.58	29.03	2017	1.65	2.12	3.26	4.33	5.79	6.61	6.53	6.27	5.22	3.87	2.57	1.86
67.13	24.9	22	4.21	4.78	6.17	7.57	7.86	7.26	6.66	5.93	6	6.08	5.06	4.03
66.98	24.8	4	3.36	3.78	4.78	5.57	6.49	5.77	4.76	4.32	4.76	4.76	3.96	3.43
73.01	31.36	184	1.57	2.38	3.57	5.52	6.85	7.52	6.09	5.54	5.14	3.99	2.59	1.81
71.43	30.2	123	1.8	2.68	4.39	6.44	8.39	10.24	8.02	7.03	6.45	4.6	2.62	1.92
68.36	26.25	38	3.03	4.07	6.02	8.28	10.58	10.96	8.96	8.16	7.36	5.85	4.03	3.01
62.75	28.81	683	2.28	3.07	4.23	5.72	7.29	8.33	8.47	7.98	6.16	4.42	3.07	2.29
70.08	33.86	1726	1.3	1.59	2.45	3.29	4.67	5.36	4.78	4.93	4.03	3.09	1.97	1.35
71.58	34.01	360	1.73	2.29	3.29	4.61	7.38	8.35	7.16	6.04	5.11	3.81	2.59	1.8
66.91	30.26	1621	1.82	2.55	3.75	5.13	7.33	8.69	8.56	7.58	6.46	4.76	3.02	1.99
72.66	32.05	188	1.79	2.61	3.95	6.31	8.42	9.05	7.34	6.26	5.64	4.47	2.69	1.82
75.61	35.3	2181	0.63	0.88	1.72	2.99	3.97	4.88	4.82	4.41	3.67	2.32	1.23	0.72
69.46	31.35	1407	1.37	1.94	3.09	4.18	5.23	6.33	5.93	5.3	4.44	3.22	1.95	1.41

Appendix A.4. TB ET_o Point Dataset Used for the Interpolation Process

References

- 1. Brasseur, G.P.; Jacob, D.; Schuck-Zöller, S. *Climate Change* 2001: Working Group II: Impacts, Adaptation and Vulnerability; Falkenmark and Lindh Quoted in UNEP/WMO; UNEP: Nairobi, Kenya, 2009.
- 2. Fangmeier, D.D.; Elliot, W.J.; Workman, S.R.; Huffman, R.L.; Schwab, G.O. *Soil and Water Conservation Engineering*, 5th ed.; Thomson: Stamford, CT, USA, 2006.
- 3. Gavilan, P.; Berengena, J.; Allen, R.G. Measuring versus estimating net radiation and soil heat flux: Impact on Penman–Monteith reference ET estimates in semiarid regions. *Agric. Water Manag.* **2007**, *89*, 275–286. [CrossRef]
- 4. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements; FAO Irrigation and Drainage Paper 56; FAO: Rome, Italy, 1998.
- López-Urrea, R.; De Santa Olalla, F.M.; Fabeiro, C.; Moratalla, A. Testing evapotranspiration equations using lysimeter observations in a semiarid climate. *Agric. Water Manag.* 2006, 85, 15–26. [CrossRef]
- 6. Allen, R.G.; Walter, I.A.; Elliott, R.L.; Howell, T.A.; Itenfisu, D.; Jensen, M.E.; Snyder, R.L. The ASCE standardised reference evapotranspiration equation. In *Task Committee on Standardization of Reference Evapotranspiration of the EWRI of the ASCE*; ASCE: Reston, VI, USA, 2005.
- 7. Başağaoğlu, H.; Chakraborty, D.; Winterle, J. Reliable Evapotranspiration Predictions with a Probabilistic Machine Learning Framework. *Water* **2021**, *13*, 557. [CrossRef]
- 8. Chakraborty, D.; Başağaoğlu, H.; Winterle, J. Interpretable vs. noninterpretable machine learning models for data-driven hydro-climatological process modeling. *Expert Syst. Appl.* **2021**, *170*, 114498. [CrossRef]
- 9. Ravindran, S.M.; Bhaskaran, S.K.M.; Ambat, S.K.N. A Deep Neural Network Architecture to Model Reference Evapotranspiration Using a Single Input Meteorological Parameter. *Environ. Process.* **2021**, *8*, 1567–1599. [CrossRef]
- Zhou, Z.; Zhao, L.; Lin, A.; Qin, W.; Lu, Y.; Li, J.; Zhong, Y.; He, L. Exploring the potential of deep factorization machine and various gradient boosting models in modeling daily reference evapotranspiration in China. *Arab. J. Geosci.* 2020, *13*, 1–20. [CrossRef]
- 11. Deo, R.C.; Wen, X.; Qi, F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl. Energy* **2016**, *168*, 568–593. [CrossRef]
- 12. Wang, L.; Kisi, O.; Zounemat-Kermani, M.; Hu, B.; Gong, W. Modeling and comparison of hourly photosynthetically active radiation in different ecosystems. *Renew. Sustain. Energy Rev.* **2015**, *56*, 436–453. [CrossRef]
- 13. Wang, L.; Kisi, O.; Zounemat-Kermani, M.; Salazar, G.; Zhu, Z.; Gong, W. Solar radiation prediction using different techniques: Model evaluation and comparison. *Renew. Sustain. Energy Rev.* **2016**, *61*, 384–397. [CrossRef]
- 14. Gocic, M.; Trajkovic, S. Software for estimating reference evapotranspiration using limited weather data. *Comput. Electron. Agric.* **2010**, *71*, 158–162. [CrossRef]
- 15. Tabari, H.; Talaee, P. Local calibration of the Hargreaves and Priestley–Taylor equations for estimating reference evapotranspiration in arid and cold climates of Iran based on the Penman–Monteith model. *J. Hydrol. Eng.* **2011**, *16*, 837–845. [CrossRef]
- Martí, P.; Royuela, A.; Manzano, J.; Palau-Salvador, G. Generalization of ETo ANN Models through Data Supplanting. J. Irrig. Drain. Eng. 2010, 136, 161–174. [CrossRef]

- 17. Rojas, J.P.; Sheffield, R.E. Evaluation of Daily Reference Evapotranspiration Methods as Compared with the ASCE-EWRI Penman-Monteith Equation Using Limited Weather Data in Northeast Louisiana. *J. Irrig. Drain. Eng.* **2013**, *139*, 285–292. [CrossRef]
- 18. Sahoo, B.; Walling, I.; Deka, B.C.; Bhatt, B.P. Standardization of Reference Evapotranspiration Models for a Subhumid Valley Rangeland in the Eastern Himalayas. *J. Irrig. Drain. Eng.* **2012**, *138*, 880–895. [CrossRef]
- 19. Shiri, J.; Nazemi, A.H.; Sadraddini, A.A.; Landeras, G.; Kisi, O.; Fard, A.F.; Marti, P. Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Comput. Electron. Agric.* **2014**, *108*, 230–241. [CrossRef]
- Ehteram, M.; Singh, V.P.; Ferdowsi, A.; Mousavi, S.F.; Farzin, S.; Karami, H.; Mohd, N.S.; Afan, H.A.; Lai, S.H.; Kisi, O.; et al. An improved model based on the support vector machine and cuckoo algorithm for simulating reference evapotranspiration. *PLoS* ONE 2019, 14, e0217499. [CrossRef]
- 21. Sayyahi, F.; Farzin, S.; Karami, H. Forecasting Daily and Monthly Reference Evapotranspiration in the Aidoghmoush Basin Using Multilayer Perceptron Coupled with Water Wave Optimization. *Complexity* **2021**, *2021*, *668375*. [CrossRef]
- 22. Tabari, H.; Talaee, P.H.; Abghari, H. Utility of coactive neuro-fuzzy inference system for pan evaporation modeling in comparison with multilayer perceptron. *Meteorol. Atmos. Phys.* **2012**, *116*, 147–154. [CrossRef]
- 23. Zakeri, M.S.; Mousavi, S.F.; Farzin, S.; Sanikhani, H. Modeling of Reference Crop Evapotranspiration in Wet and Dry Climates Using Data-Mining Methods and Empirical Equations. *J. Soft Comput. Civ. Eng.* **2022**, *6*, 1–28. [CrossRef]
- 24. Fooladmand, H.R.; Zandilak, H.; Ravanan, M.H. Comparison of different types of Hargreaves equation for estimating monthly evapotranspiration in the south of Iran. *Arch. Agron. Soil Sci.* 2008, 54, 321–330. [CrossRef]
- 25. George, B.A.; Reddy, B.R.S.; Raghuwanshi, N.S.; Wallender, W.W. Decision support system for estimating reference evapotranspiration. J. Irrig. Drain. Eng. 2002, 128, 1–10. [CrossRef]
- Sabziparvar, A.A.; Tabari, H. Regional estimation of reference evapotranspiration in arid and semiarid regions. J. Irrig. Drain. Eng. 2010, 136, 724–731. [CrossRef]
- Tabari, H. Evaluation of reference crop evapotranspiration equations in various climates. *Water Resour. Manag.* 2010, 24, 2311–2337. [CrossRef]
- Xu, C.Y.; Singh, V.P. Cross comparison of empirical equations for calculating potential evapotranspiration with data from Switzerland. *Water Resour. Manag* 2002, 16, 197–219. [CrossRef]
- Anaraki, M.V.; Farzin, S.; Mousavi, S.-F.; Karami, H. Uncertainty Analysis of Climate Change Impacts on Flood Frequency by Using Hybrid Machine Learning Methods. *Water Resour. Manag.* 2021, 35, 199–223. [CrossRef]
- 30. Farzin, S.; Anaraki, M.V. Modeling and predicting suspended sediment load under climate change conditions: A new hybridization strategy. *J. Water Clim. Chang.* **2021**, *12*, 2422–2443. [CrossRef]
- 31. Kumar, M.; Bandyopadhyay, A.; Raghuwanshi, N.S.; Singh, R. Comparative study of conventional and artificial neural networkbased ETo estimation models. *Irrig. Sci.* 2008, *26*, 531–545. [CrossRef]
- 32. Kumar, M.; Raghuwanshi, N.S.; Singh, R. Artificial neural networks approach in evapotranspiration modeling: A review. *Irrig. Sci.* 2010, *29*, 11–25. [CrossRef]
- Landeras, G.; Ortiz-Barredo, A.; López, J.J. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). *Agric. Water Manag.* 2008, 95, 553–565. [CrossRef]
- 34. Khoob, A.R. Comparative study of Hargreaves's and artificial neural network's methodologies in estimating reference evapotranspiration in a semiarid environment. *Irrig. Sci.* 2007, *26*, 253–259. [CrossRef]
- 35. Chia, M.Y.; Huang, Y.F.; Koo, C.H.; Fung, K.F. Recent Advances in Evapotranspiration Estimation Using Artificial Intelligence Approaches with a Focus on Hybridisation Techniques—A Review. *Agronomy* **2020**, *10*, 101. [CrossRef]
- Yin, Z.; Feng, Q.; Yang, L.; Deo, R.C.; Wen, X.; Si, J.; Xiao, S. Future Projection with an Extreme-Learning Machine and Support Vector Regression of Reference Evapotranspiration in a Mountainous Inland Watershed in North-West China. *Water* 2017, *9*, 880. [CrossRef]
- 37. Wen, X.; Si, J.; He, Z.; Wu, J.; Shao, H.; Yu, H. Support-Vector-Machine-Based Models for Modeling Daily Reference Evapotranspiration With Limited Climatic Data in Extreme Arid Regions. *Water Resour. Manag.* **2015**, *29*, 3195–3209. [CrossRef]
- Wang, S.; Fu, Z.-Y.; Chen, H.; Nie, Y.-P.; Wang, K.-L. Modeling daily reference ET in the karst area of northwest Guangxi (China) using gene expression programming (GEP) and artificial neural network (ANN). *Theor. Appl. Climatol.* 2016, 126, 493–504. [CrossRef]
- 39. Sanikhani, H.; Kisi, O.; Maroufpoor, E.; Yaseen, Z.M. Temperature-based modeling of reference evapotranspiration using several artificial intelligence models: Application of different modeling scenarios. *Theor. Appl. Climatol.* **2019**, *135*, 449–462. [CrossRef]
- 40. Pour, O.M.R.; Piri, J.; Kisi, O. Comparison of SVM, ANFIS and GEP in modeling monthly potential evapotranspiration in an arid region (Case study: Sistan and Baluchestan Province, Iran). *Water Supply* **2019**, *19*, 392–403. [CrossRef]
- 41. Wu, L.; Peng, Y.; Fan, J.; Wang, Y. Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. *Hydrol. Res.* **2019**, *50*, 1730–1750. [CrossRef]
- Saggi, M.K.; Jain, S. Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning. Comput. Electron. Agric. 2019, 156, 387–398. [CrossRef]

- 43. Shiri, J.; Nazemi, A.H.; Sadraddini, A.A.; Marti, P.; Fard, A.F.; Kisi, O.; Landeras, G. Alternative heuristics equations to the Priestley–Taylor approach: Assessing reference evapotranspiration estimation. *Appl. Clim.* **2019**, *138*, 831–848. [CrossRef]
- Tikhamarine, Y.; Malik, A.; Kumar, A.; Souag-Gamane, D.; Kisi, O. Estimation of monthly reference evapotranspiration using novel hybrid machine learning approaches. *Hydrol. Sci. J.* 2019, 64, 1824–1842. [CrossRef]
- 45. Ferreira, L.B.; da Cunha, F.F.; de Oliveira, R.A.; Filho, E.I.F. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM—A new approach. *J. Hydrol.* **2019**, *572*, 556–570. [CrossRef]
- Granata, F. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agric. Water Manag.* 2019, 217, 303–315. [CrossRef]
- Keshtegar, B.; Kisi, O.; Zounemat-Kermani, M. Polynomial chaos expansion and response surface method for nonlinear modelling of reference evapotranspiration. *Hydrol. Sci. J.* 2019, *64*, 720–730. [CrossRef]
- Nourani, V.; Elkiran, G.; Abdullahi, J. Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. J. Hydrol. 2019, 577, 123958. [CrossRef]
- 49. Shiri, J. Modeling reference evapotranspiration in island environments: Assessing the practical implications. *J. Hydrol.* **2019**, 570, 265–280. [CrossRef]
- Raza, A.; Hu, Y.; Shoaib, M.; Elnabi, M.K.A.; Zubair, M.; Nauman, M.; Syed, N.R. A Systematic Review on Estimation of Reference Evapotranspiration under Prisma Guidelines. *Pol. J. Environ. Stud.* 2021, 30, 5413–5422. [CrossRef]
- 51. Smith, M.; Allen, R.G.; Monteith, J.L.; Pereira, L.S.; Perrier, A.; Pruitt, W.O. Report on the Expert Consultation on Revision of FAO Guidelines for Crop Water Requirements, Land and Water Development; Division, FAO: Rome, Italy, 1991.
- Sarfaraz, S.; Arsalan, M.H.; Fatima, H. sRegionalising the climate of Pakistan using Köppen classification system. *Pak. Geogr. Rev.* 2014, 69, 111–132. Available online: http://www.academia.edu/download/39532009/PGR_2014_Vol_69_No_02_article_05.pdf (accessed on 25 March 2022).
- Raza, A.; Shoaib, M.; Khan, A.; Baig, F.; Faiz, M.A.; Khan, M.M. Application of non-conventional soft computing approaches for estimation of reference evapotranspiration in various climatic regions. *Theor. Appl. Climatol.* 2020, 139, 1459–1477. [CrossRef]
- Raza, A.; Shoaib, M.; Faiz, M.A.; Baig, F.; Khan, M.M.; Ullah, M.K.; Zubair, M. Comparative Assessment of Reference Evapotranspiration Estimation Using Conventional Method and Machine Learning Algorithms in Four Climatic Regions. J. Pure Appl. Geophys. 2020, 177, 4479–4508. [CrossRef]
- Shoaib, M. Impact of Wavelet Transformation on Data Driven Rainfall-Runoff Models. Ph.D. Thesis, University of Auckland, Auckland, New Zealand, 2016. Available online: http://hdl.handle.net/2292/28463 (accessed on 22 March 2022).
- Shoaib, M.; Shamseldin, A.Y.; Melville, B.W.; Khan, M.M. Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach. J. Hydrol. 2015, 527, 326–344. [CrossRef]
- 57. Martí, P.; Manzano, J.; Royuela, A. Assessment of a 4-input artificial neural network for ET o estimation through data set scanning procedures. *Irrig. Sci.* 2011, *29*, 181–195.
- 58. Wang, K.; Sun, J.; Cheng, G.; Jiang, H. Effect of altitude and latitude on surface air temperature across the Qinghai-Tibet Plateau. *J. Mt. Sci.* **2011**, *8*, 808–816. [CrossRef]
- 59. Turgut, E.T.; Usanmaz, Ö. An analysis of vertical profiles of wind and humidity based on long-term radiosonde data in Turkey. *Anadolu Univ. J. Sci. Technol. A-Appl. Sci. Eng.* **2016**, *17*, 830–844.
- 60. Droogers, P.; Allen, R.G. Estimating Reference Evapotranspiration Under Inaccurate Data Conditions. *Irrig. Drain. Syst.* 2002, *16*, 33–45. [CrossRef]
- 61. Kisi, O.; Sanikhani, H.; Zounemat-Kermani, M.; Niazi, F. Long-term monthly evapotranspiration modeling by several data-driven methods without climatic data. *Comput. Electron. Agric.* **2015**, *115*, 66–77. [CrossRef]
- 62. Martí, P.; González-Altozano, P.; Gasque, M. Reference evapotranspiration estimation without local climatic data. *Irrig. Sci.* 2011, 29, 479–495. [CrossRef]
- 63. Zhu, B.; Feng, Y.; Gong, D.; Jiang, S.; Zhao, L.; Cui, N. Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. *Comput. Electron. Agric.* 2020, 173, 105430. [CrossRef]